

Development of a Realistic Three Dimensional Structure for Bovine Type I Collagen Using Computational Molecular Modeling Methods

Three-dimensional structures of proteins are important in defining structure-function relationships. Most such structures are obtained from X-ray crystallography. Because collagen has not been crystallized, computerized molecular modeling methods were applied to the development of three-dimensional structures for the triple helices and 'Smith' microfibril of Type I collagen. The initial model was based on a triple helix of (Gly-Pro-Hyp)₁₂. Amino acid sequences of the alpha 1 and 2 chains of calf skin Type I collagen were substituted into the (Gly-Pro-Hyp)₁₂ model. Each Type I structure was refined with the Kollman force field energy minimization calculation to form all possible stabilizing interactions within the protein. Specific interaction sites which may contribute to the stability of the collagen molecule can be identified in the triple helical and microfibril models. When the microfibril model was extended to a full D-space, a set of bands was observed which reproduced the negative and positive stained bands observed in electron micrographs of Type I collagen fibers. These models may be used to target potential domains in collagen that contain sites for crosslinking reactions with chromium and other agents; determine what constraints on size, shape and molecular characteristics are likely to describe the optimum crosslinking agents; and suggest the application of which analytical and spectroscopic techniques will give the most rigorous test of the computer models.

Resumen

Las estructuras tri-dimensionales de proteínas son importantes para definir relaciones entre estructura y función. La mayoría de estas estructuras son obtenidas por cristalografía de rayos X. Puesto que el colágeno no ha sido cristalizado, métodos computarizados de modelaje molecular fueron aplicados hacia el desarrollo de estructuras tri-dimensionales para las hélices triplicadas y las microfibrillas 'Smith' del colágeno tipo I. El modelo inicial fue basado en la hélice triple de (Gly-Pro-Hyp)₁₂. Secuencias de amino ácidos en las cadenas 1 y 2 de colágeno tipo I en piel vacuna fueron substituidas en el modelo (Gly-Pro-Hyp)₁₂. Cada estructura tipo I fue refinada con el cálculo de minimización del campo de energía de fuerza Kollman para formar toda posible interacción estabilizadora dentro de la proteína. Sitios específicos de interacción los cuales pueden contribuir en la estabilidad de la molécula de colágeno pueden ser identificados en los modelos de triple hélice y microfibrillas. Cuando el modelo de la microfibrilla fue extendido a un espacio-D lleno u ocupado, se observó un set de bandas las cuales reprodujeron bandas negativa y positivamente tiñadas vistas en las micrográficas de electrones en las fi-

bras de colágeno tipo I. Estos modelos pueden ser utilizados para afrontar dominios potenciales en el colágeno que contenga sitios para reacciones de enlace cruzado con cromo y otros agentes; determinar que constringe el tamaño, forma y características moleculares son índices que describen los agentes óptimos de enlace cruzado; y sugieren la aplicación de técnicas analíticas y espectroscópicas que proveeran la más rigurosa prueba de los modelos de computadora.

Zusammenfassung

Dreidimensionale Proteinstrukturen sind für die Definition der Struktur-Funktionsbeziehungen wichtig. Die meisten solcher Strukturen liefert die Röntgenkristallographie. Da Kollagen bisher nicht kristallisiert vorliegt, wurden computerunterstützte Methoden der Molekülmodellierung zur Entwicklung dreidimensionaler Strukturen der Tripelhelices und 'Smith'-Mikrofibrillen des Typ I Kollagen verwendet. Das Ausgangsmodell basierte auf der Tripelhelix von (Gly-Pro-Hyp)₁₂. Aminosäuresequenzen der Alpha 1 und 2 Ketten von Typ I Kalbshautkollagen wurden in das (Gly-Pro-Hyp)₁₂-Modell eingebaut. Jede Typ I Struktur wurde mit der Kollmannschen Kraftfeldenergieminimierung verfeinert, um so alle möglichen stabilisierenden Wechselwirkungen innerhalb des Proteins zu erfassen. Spezifische Wechselwirkungsstellen, welche zur Stabilisierung des Kollagenmoleküls beitragen könnten, können in den Modellen für Tripelhelices und Mikrofibrillen festgestellt werden. Wenn das Mikrofibrillenmodell auf einen vollen dreidimensionalen Raum erweitert wird, ist ein Satz von Bändern zu erkennen, der die in elektronenmikroskopischen Aufnahmen der Typ I Kollagenfasern zu beobachtenden negativ und positiv gefärbten Banden wiedergibt. Diese Modelle können dazu benutzt werden, mögliche Domänen in Kollagen, die Stellen für Quervernetzungsreaktionen mit Chrom oder anderen Agentien enthalten, ins Ziel zu nehmen. Darüberhinaus kann ermittelt werden, welche Einschränkungen von Größe, Gestalt und molekularen Charakteristika geeignet sind, die optimalen Quervernetzungsreagentien zu beschreiben. Fernerhin kann man daraus eine Vorstellung gewinnen, welche analytischen und spektroskopischen Techniken für die strengste Prüfungen der Computermodelle angewendet werden sollten.

Development of a realistic three dimensional structure for bovine type I collagen using computational molecular modeling methods

E. M. Brown, J. M. Chen and S. H. Fairheller
Eastern Regional Research Center, USDA, Philadelphia,
PA 19118

Abstract

Three-dimensional structures of proteins are important in defining structure-function relationships. Most such structures are obtained from X-ray crystallography. Because collagen has not been crystallized, computerized molecular modeling methods were applied to the development of three-dimensional structures for the triple helices and "Smith" microfibril of Type I collagen. The initial model was based on a triple helix of (Gly-Pro-Hyp)₁₂. Amino acid sequences of the alpha 1 and 2 chains of calf skin

Type I collagen were substituted into the (Gly-Pro-Hyp)₁₂ model. Each Type I structure was refined with the Kollman force field energy minimization calculation to form all possible stabilizing interactions within the protein. Specific interaction sites which may contribute to the stability of the collagen molecule can be identified in the triple helical and microfibril models. These models may be used to target potential domains in collagen that contain sites for crosslinking reactions with chromium and other agents; determine the constraints on size, shape and molecular characteristics that are likely to describe the optimum crosslinking agents; and suggest the application of analytical and spectroscopic techniques that will give the most rigorous test of the computer models.

Introduction

Effective crosslinking of collagen in animal hides produces leather, having properties of strength and resistance to organisms that would attack and destroy the hide. At present, Cr(III) is the most commonly used tanning agent. Although other materials (vegetable tannins, synthetic organic crosslinkers and mineral agents) are used in specific tannages, trivalent chromium is the most effective in producing high quality leather from calf skin⁽¹⁾. Concerns over possible restrictions on the availability, use and disposal of chromium-containing materials have prompted research into the development of other tannages⁽²⁻⁴⁾. Any new agent must have tanning properties equal or superior to that of Cr(III) and be without undesirable attributes. At present tanning is more art than science; procedures for identifying alternative agents are based on the empirical testing of compounds or mixtures having properties similar to those of known tanning agents⁽⁵⁻⁸⁾.

An understanding of the stabilizing interactions of native collagen and the identification of reaction sites for known tanning ligands in the tertiary structure of collagen are important for the rational design of new, specific crosslinking reagents. A model of Type I collagen combined with structure-function studies based on protein-ligand interactions would make it possible to determine the potential chemical and geometric requirements for tanning ligands, which will in turn lead to more efficient and effective procedures for designing new or improved tanning agents. A key barrier to this approach is the lack of a three-dimensional structure of Type I collagen. This study describes the application of molecular modeling to the study of its tertiary structure.

Structurally, this semi-flexible, rod-like molecule is approximately 300 nm in length and 1.0 – 1.4 nm in diameter, depending upon the hydrated state of the protein. The basic unit of Type I collagen is a right-handed triple helix composed of two α -1 and one α -2 chains, each with a left-handed helical conformation. Excluding the extra-helical amino and carboxyl terminal telopeptides, there are 1014 amino acid residues per polypeptide chain, such that the consensus sequence of Gly-X-Y is repeated 338 times⁽⁹⁾. The sequences of the α -1 and α -2 chains are similar, each containing 33% glycine and 25% imino acids (proline and hydroxyproline)⁽¹⁰⁾. Approximately 3.3 amino acid residues are necessary to make one complete (360 degrees) rotation about the polypeptide helical long axis⁽¹¹⁾.

The axial arrangement of collagen molecules within fibrils has been studied extensively by electron microscopy and X-ray diffraction and is known to be highly ordered^(9,12). The banding patterns observed in electron micrographs of positively and negatively stained collagen can be attributed^(13,14) to an array of five triple helical domains adjacently aligned with a 1 D interval stagger (Figure 1) where 4.4 D intervals constitute the entire length

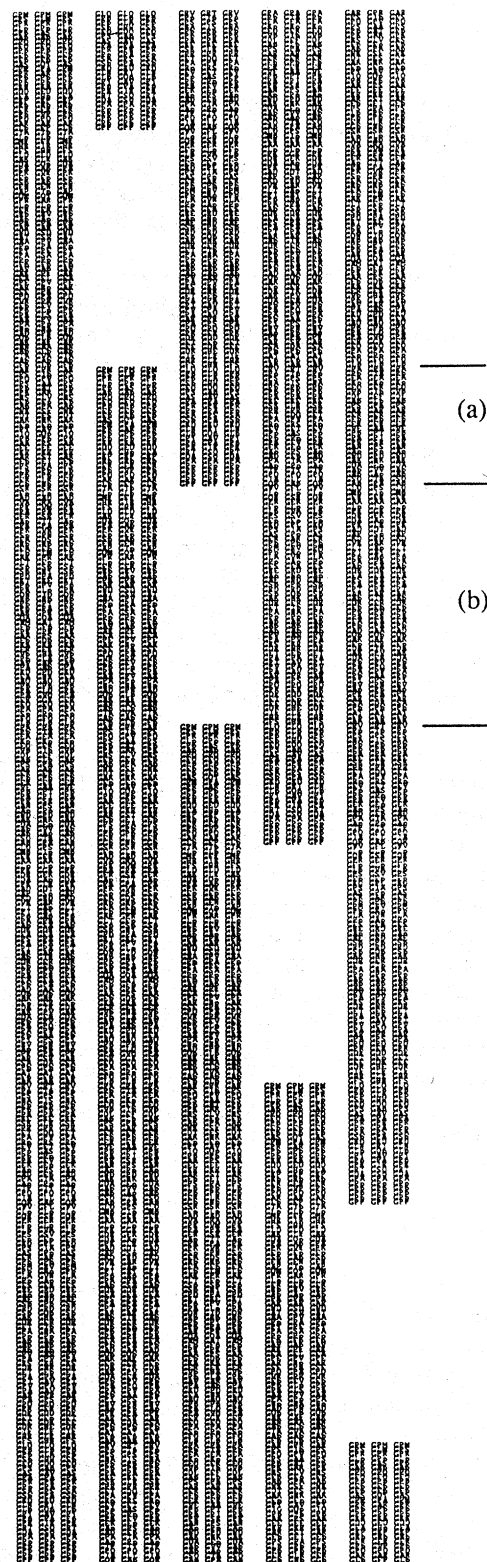


Figure 1. The amino acid sequence of bovine Type I collagen (greatly reduced) and arranged to show the alignment of polypeptide chains in the complete microfibril. The design regions are: (a) an overlap 0.33 D space in length; (b) gap region 0.67 D space. One complete D space is equal to the sum of (a) and (b); the entire molecule is 4.4 D space in length

of a single collagen molecule⁽⁹⁾. A "gap" region corresponding to a length of 0.6 D⁽⁹⁾ separates individual collagen molecules end-to-end. Smith⁽¹⁴⁾ incorporated these data into a three-dimensional packing model where five collagen triple helices are packed in a circular array or microfibril exhibiting an overall left-handed supercoil.

The Smith microfibril is postulated to be an intermediate stage in fibril formation^(15,17). At the fibril level, the dense packing of collagen molecules may result in more liquid-like properties near the center with crystalline features becoming more apparent near the periphery of the fibril^(12,18). Other models proposed for the microfibril and fibril include: hexagonal and 'quasi'-hexagonal packing models⁽¹⁹⁻²¹⁾; non-crystalline, more liquid-like lateral packing models^(22,23); and models such as the octafibril and the simple two-dimensional liquid model^(22,24) that regard the fibril as having no intermediate substructure.

We have used molecular modeling techniques to construct a three-dimensional model for a fragment of the 'Smith' microfibril of Type I collagen. With this model we are attempting to identify sidechain-sidechain and sidechain-backbone interactions between amino acid residues that may contribute to intra- or inter-helical stabilization and to visualize the effects of processing conditions on potential sites for tanning ligands.

Methods

Molecular modeling studies were conducted on a Unix based high resolution graphics workstation using a commercial molecular modeling program. This color graphics system allowed for detailed visualization and real time manipulation of chemical structures. The molecular modeling program contains a dictionary of average geometric parameters derived from X-ray crystallographic data for each amino acid residue. The program also contains algorithms and force fields for structural optimization through energy minimization. The force field or potential energy function is given as a sum of bond energies and non-bonded interaction terms. The algorithm, Amber (Assisted Model Building and Energy Refinement), uses the Kollman^(25,26) force field to calculate the difference in energy between an ideal geometry and the actual geometry of the molecule. Amber allows for the calculation of hydrogen bond energies and is considered to be most accurate for peptides and proteins⁽²⁷⁾. Parameters used with Amber include atomic partial charges as calculated by Kollman using a united atom approach with only essential hydrogens.

The calculated energy of the model was minimized first with the Simplex⁽²⁸⁾ method, a nonderivative calculation that moves one atom at a time, to achieve a "good" geometry, and then with a conjugate gradient method to further refine the model. Solvent molecules were not explicitly included in the models, but a distance dependent dielectric function was used to implicitly account for the effect of solvent. All minimizations were carried out *in vacuo*, using a root-mean square (rms) derivative of 0.01 kcal/mole as a cutoff value for ending the minimization.

The initial step⁽²⁹⁾ in the construction of the Type I 'Smith' microfibril model was to build single chains of (Gly-Pro-Pro)₄ and (Gly-Pro-Hyp)₄ using the parameters provided by the dictionary component of the molecular modeling program. Torsional angles assigned for the peptide backbone were those reported by Miller and Scheraga⁽³⁰⁾ for the lowest energy structure for the triple helix of (Gly-Pro-Pro)₄ as obtained from ECEPP (Empirical Conformational Energy for Proteins and Peptides). An N-acetyl blocking group was added to the amino terminus and an N-methylamide group to the carboxyl terminus of each polypeptide chain to ac-

count for end effects from neighboring residues. These structures were refined using the Kollman force field with Amber^(25,26). Deviations between initial and relaxed structures (rms < 0.8 Å for all atoms) showed that the chosen angles were suitable. Three energy refined polypeptide chains were interactively docked so that the resulting structure was a triple helix.

Once this unit structure, the triple helix 3(Gly-X-Y)₄, was refined through energy minimization, the helical units were docked carboxyl to amino terminus to extend the structure to the (Gly-X-Y)₁₂ level. To do this, all N- and C-terminal blocking groups were first removed and the triple-helical segments aligned so that the N-terminal ends were adjacent to the C-terminal ends. Peptide bonds were formed between neighboring polypeptide chains, thereby linking together the triple-helical segments. Prior to refinement of the final structure, N-acetyl and N-methylamide blocking groups were reassigned to the respective ends of each open polypeptide chain. For collagen, the energy refinement of a large structure consisting of smaller previously energy minimized structures was much more efficient than the refinement of a large unminimized structure.

The microfibril was constructed⁽²⁹⁾ by docking five triple helices in accordance with the Smith⁽¹⁴⁾ model. The initial stage in this process was the packing of two triple-helical segments of 3(Gly-Pro-Pro)₁₂, a choice which allowed us to utilize the results of Nemethy and Scheraga⁽³⁰⁾ for the packing of two (Gly-Pro-Pro)₅ triple helices. The dimer was then refined through energy minimization. The same procedure was followed for the packing of the third, fourth and fifth 3 (Gly-Pro-Pro)₁₂ helix. All atoms of each structure were constrained to their original positions during docking and energy minimization. This step insured that the best possible packing interactions would occur between molecules in their starting conformations. These constraints were removed after the packing positions were optimized for each substructure polypeptide chain, and the completed structure was then subjected to further energy refinement. Once the microfibril was built, portions of the amino acid sequences of the α -1 and α -2 chains of Type I collagen were substituted into the microfibril model^(31,32) and the structure was again refined by energy minimization. The specific microfibril region in Table 1 was chosen for study because it contains regions with many charged sidechains as well as a hydrophobic patch.

Results and Discussion

In the present model, the sequences of adjacent triple helical segments (C1 - C5) were offset by 234 residues (one D-spacing) (Table 1). Each triple helix has three-fold rotational symmetry about its helical axis and the microfibril has five-fold rotational symmetry about its longitudinal axis. A cross section of this fragment of a 'Smith' microfibril model is primarily a circular array of five triple-helical structures with a short 'gap' region containing four helical structures. Each triple helix has a left-handed twist around the microfibril structure and the tilt angle of each molecule is approximately 10 degrees with respect to the microfibrillar long axis⁽²⁹⁾. The collagen polypeptide chain, triple helix and microfibril have a left, right and left handed twist, respectively. This process of alternating the twist handedness at each higher level of fibril formation probably maximizes fibrillar strength⁽³³⁾.

In the space-filling models of the energy minimized triple helical structures (Figure 2) and 'Smith' microfibril (Figure 3) the carboxyl terminus, tripeptide 338, is at the upper end of C2. All five triple helical structures are aligned in the same direction and the amino acid sequence defines several interesting functional

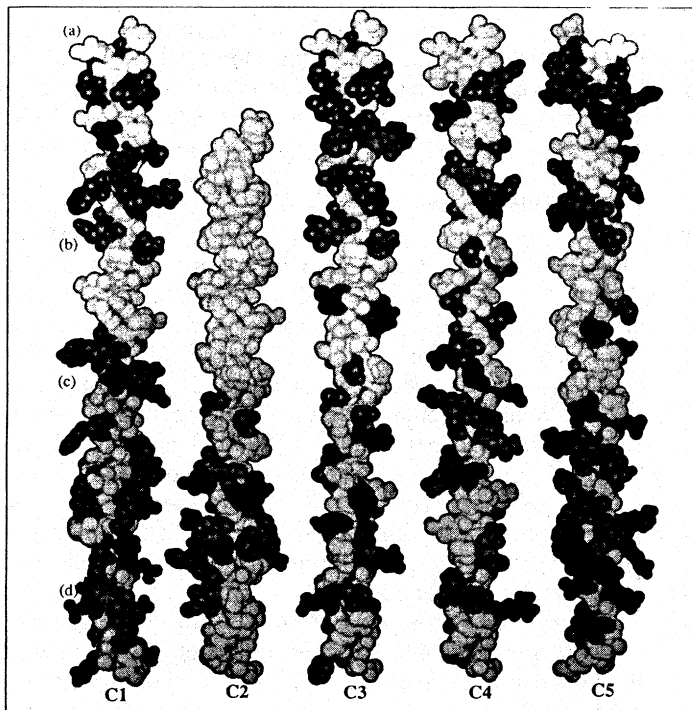


Figure 2. Space-filling models for the C1 – C5 (Table 1) triple helical fragments of Type I collagen. Amino terminal ends of the fragments are at the bottom in this representation. Four classes of sidechains are distinguished by shading from very light to dark (a) structural (Gly, Pro, Hyp), (b) nonpolar, (c) basic, (d) acidic

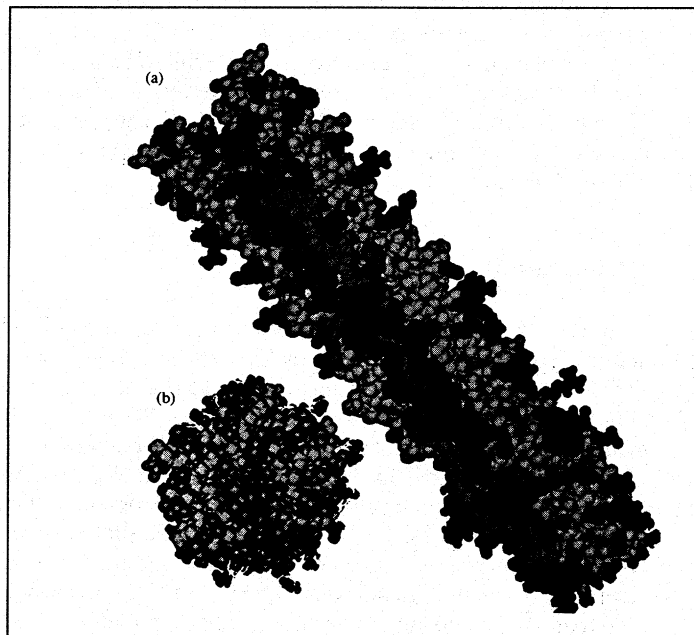


Figure 3. Lateral view (a) of the space-filling model for the microfibril constructed from the triple helical fragments of Figure 2. Shading is as described for Figure 2 except that the polypeptide chains of C2 are shaded individually to emphasize the path of this fragment across the lateral view of the microfibril ending in the gap region. Cross sectional view (b) of the microfibril fragment

domains. For example, non-polar sidechains tend to be found near the top of each triple helical model and acidic or basic sidechains near the bottom in this selection of fragments.⁴ The clustering of functional groups is evident in both lateral and cross sectional views of the Type I microfibril model (Figure 3) and may indicate those types of interactions important for fibril stability. The cross sectional view of the microfibril model (Figure upper-right) shows the positioning of non-polar and charged sidechains around the circumference of the structure.

Energetic Evaluation of the Type I Collagen Models

The computed total potential energies of stabilization (ΔE after minimization, for the $(\text{Gly-Pro-Pro})_{12}$ and $(\text{Gly-Pro-Hyp})_{12}$ triple helices were -554 and -572 kcal/mole⁽²⁹⁾, respectively. In contrast, (ΔE_{tot}) for the individual Type I triple helical segments C1, C3, C4, and C5 (Table 1) were between -970 and -1235 kcal/mole⁽³¹⁾. The carboxyl terminal C2 is both shorter and less stable as a triple helix ($\Delta E_{\text{tot}} = -404$ kcal/mole). At the next level, the Type I microfibril segment ($\Delta E_{\text{tot}} = -6004$ kcal/mole) was significantly more stable than microfibrils composed of $(\text{Gly-Pro-Pro})_{12}$ or $(\text{Gly-Pro-Hyp})_{12}$ ($\Delta E_{\text{tot}} = -550, -1041$ kcal/mole, respectively)⁽⁶⁾. Although a direct comparison of the potential energy values cannot be made for structures containing different amino acid sequences, the large difference observed when comparing the Type I model to the $(\text{Gly-Pro-Hyp})_{12}$ model indicates that sidechain interactions are important for stabilizing both triple helical and microfibrillar structures of collagen.

Analysis of Sidechain Interactions in the Stabilization of the Model Structure

Arginine sidechains are important in defining the tertiary structure of collagen because they are positively charged, bulky, and have the five amino protons (potential hydrogen bond donors). In the energy minimized triple helical models (Figure 1), arginine sidechains in X positions of Gly-X-Y sequences appear to be folded towards the helical axis while those of arginine in Y positions extend outward from the helical long axis. In the microfibril model (Figure 3), arginine side chains in Y positions make an angle of $116^\circ \pm 11^\circ$ with the helical axis, while those in X positions make an angle of $71^\circ \pm 12^\circ$ ⁽³⁴⁾. This difference suggests that arginine in the X position may stabilize triple helical structure by reducing the conformational flexibility of the polypeptide chain while in the Y position it may function in inter-helical packing where the sidechains form ion-pairs with acidic groups on adjacent helices. In calf skin Type I collagen, 82% of the arginine residues are in Y positions, suggesting a role for arginine in collagen packing.

Triple helical Type I collagen contains 63 tripeptide sequences of the type Gly-X-Lys and 30 of the type Gly-Lys-Y⁽⁹⁻¹¹⁾. The close proximity of acidic groups (aspartate and glutamate) to the lysine-rich regions allows ion-pair formation to stabilize collagen structure. In these models both intra- and inter-polypeptide interactions are observed. The stabilizing effects of intra-polypeptide aspartate-glutamate interactions were reported by Katz and David⁽⁶⁾ in their study, however, did not include interactions between triple helices.

Non-polar sidechains (leucine, isoleucine, valine, phenylalanine, methionine, proline and alanine) contribute to the packing of collagen molecules through hydrophobic interactions. The non-polar sidechains are localized in clusters along the helical long axis of the collagen polypeptide chains. Figure 2 shows that non-polar sidechains (b) are directed outward, away from the center of the triple helical models. The explicit inclusion of solvent (water) in the models would result in unfavorable free-energies

hydration, as the non-polar surfaces of collagen triple helices would tend to be packed together⁽³⁶⁾, and shielded from the hydrophilic environment. Hence, these non-polar sidechains found in clusters along the Type I sequences contribute to the hydrophobic helical packing forces between different collagen molecules^(33,37).

In addition to ion-pair formation with basic residues, acidic sidechains (aspartate, glutamate) may be hydrogen-bonded to the hydroxyl groups of serine, threonine and hydroxyproline or the amine groups of asparagine and glutamine. The carboxy terminal triple helix (C4) contains a set of hydrogen bonded asparagine-aspartate sidechains that help to stabilize the three-dimensional model. In the microfibrillar model (Figure 3) hydroxyprolines from adjacent collagen molecules are clustered in the interior of the microfibril where their hydroxyl groups tend to be oriented toward acidic sidechains.

Evaluation of Potential Sites for Ligand Binding

The acidic residues aspartate and glutamate provide potential binding sites for Cr(III)⁽⁹⁾. Crosslinking via a binuclear chrome complex requires the distance between acid groups on different triple helices to be 6 to 8 Å⁽³⁸⁾. In the native form of this microfibril fragment, eight such acidic sites were identified, and six more after hydrolysis of glutamine and asparagine side chains to the corresponding acid form by liming⁽³⁹⁾. Four sites, involving three triple helices each, suitable for a trinuclear Cr(III) complex⁽³⁸⁾ were identified.

In order to react with and successfully stabilize the collagen matrix, Cr(III) or other acidic site tanning agents must have a higher affinity/activity than lysine or arginine for the acidic groups of collagen. A side effect of the reaction of Cr(III) with acidic groups on collagen is that basic sidechains no longer ion-paired with acidic groups would then be available for cross-linking with other modifying agents. Hence Cr(III) could not only stabilize collagen structure but also create possible new sites for further stabilization by basic site reagents.

Although ion pairs between the ε-amino group of lysine and acidic side chains can provide both intra- and inter- triple helical stabilization to the collagen molecule, the flexibility of lysine side chains within proteins⁽⁴⁰⁾ suggests that these interactions are less specific than those of arginine residues. As a primary amine, the ε-amino group of lysine makes an attractive target for acidic crosslinking reagents. Scholnick et al.⁽⁴¹⁾ effectively crosslinked sheepskin with esters of 8 to 12-carbon dicarboxylic acids. In the microfibril model, 10 potential binding sites for methylene chains 8 and 10 Å apart were identified. The inability of shorter chain acids to form effective crosslinks was explained by a lack of binding sites at smaller distances involving more than one triple helix.

In order to design more effective tanning agents, reagents that can compete better than Cr(III) for the acidic sites must be identified. Computationally, 'Thermodynamic Cyclic-Perturbation Methods'⁽⁴²⁻⁴⁴⁾ could be used to calculate and compare relative free energies of binding for different potential tanning ligands. To determine those geometric parameters that result in the most effective distribution of the tanning agent throughout the collagen matrix, the three dimensional structure of collagen packing and dynamics is being studied. Both the geometric and chemical properties that define a specific and ideal tanning ligand must be incorporated into the design of the new ligand.

References

1. Bienkiewicz, K. J. (1983) in 'Physical Chemistry of Leather Making', R. E. Krieger Pub. Co., Florida, p 324
2. Das Gupta, S. (1986) *Leather Science* 33:2
3. Evans, N. A., Milligan, B. & Montgomery, K. C. (1987) *J. Am. Leather Chem. Assoc.* 82:86
4. Ying, L., Scholnick, F. & Fairheller, S. H. (1990) *J. Am. Leather Chem. Assoc.* 85:78.
5. Deme, I. (1989) *Proceedings XXth Congress of the IULTCS*, #S-4
6. He, X., & Guo, Z. (1989) *Proceedings XXth Congress of the IULTCS*, #S-11
7. Heidemann, E., Tonigold, L., & Hein, A. (1989) *Proc. XXth Congress IULTCS*, #S-12
8. Pauckner, W. (1989) *Proc. of the XXth Congress IULTCS*, #S-14
9. Piez, K. A. (1984) in 'Extracellular Matrix Biochemistry' Piez, K. A., Reddi, A. H. (eds.), Elsevier, N. Y., p 1
10. Ramachandran, G. N. & Ramakrishnan, C. (1976) in 'Biochemistry of Collagen'. Ramachandran, G. N. (ed.). Plenum Press, New York. p 45
11. Ramachandran, G. N. (1968) *J. Am. Leather Chem. Assoc.* 63:160
12. Chapman, J. A. (1984) in 'Connective Tissues Matrix: Topics in Molecular and Structural Biology' Vol. 5, Hukins, D. W. L. (ed.). Verlag Chemie, Weinheim. p 89
13. Van der Rest, M. & Garrone, R. (1990) *Biochimie* 72:473
14. Smith, J. W. (1968) *Nature* 219:157
15. Veis, A., Miller, A., Leibovich, S. J. & Traub, W. (1979) *Biochim. Biophys. Acta* 576:88
16. Na, G. C., Butz, L. J. & Carroll, R. J. (1986) *J. Biol. Chem.* 261:12290
17. Na, G. C., Butz, L. J., Bailey, D. G. & Carroll, R. J. (1986) *Biochemistry* 25:958
18. Hulmes, D. J. S., Jesior, J.-C., Miller, A., Berthet-Colominas, C. & Wolff, C. (1981) *Proc. Natl. Acad. Sci. USA* 78:3567
19. Miller, A., & Wray, J. S. (1971) *Nature* 230:437
20. Hulmes, D. J. S., & Miller, A. (1979) *Nature* 282:878
21. Miller, A. (1982) *Trends Biochem. Sci.* 7:13
22. Woodhead-Galloway, J. (1984) in 'Connective Tissue Matrix: Topics in Molecular and Structural Biology' Vol. 5, Hukins, D. W. L. (ed.). Verlag Chemie, Weinheim. p 133
23. Giraud-Guille, M.-M. (1987) *Mol. Cryst. Liq. Cryst.* 153:15
24. Hosemann, R., Dressig, W. & Nemetschek, T. (1974) *J. Mol. Biol.* 83:275
25. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. A. (1984) *J. Am. Chem. Soc.* 106:765
26. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comput. Chem.* 7:230. 30
27. Hall, D. & Pavitt, N. (1984) *J. Comput. Chem.* 5:441
28. Sybyl (1990) Macromolecular modeling software, version 5.32, Tripos Associates, Inc
29. Chen, J. M., Kung, C. E., Fairheller, S. H., & Brown, E. M. (1991) *J. Protein Chem.* 10:535
30. Nemethy, G. & Scheraga, H. A. (1984) *Biopolymers* 23:2781
31. Chen, J. M., Fairheller, S. H., & Brown, E. M. (1991) *J. Am. Leather Chem. Assoc.* 86:475
32. Chen, J. M., Fairheller, S. H., & Brown, E. M. (1991) *J. Am. Leather Chem. Assoc.* 86:487
33. Piez, K. A. & Trus, B. L. (1977) *J. Mol. Biol.* 110:701
34. Brown, E. M., Chen, J. M. & Fairheller, S. H. (1992) *J. Am. Leather Chem. Assoc.* (In press)
35. Katz, E. P. & David, C. W. (1990) *Biopolymers* 29:791
36. Piez, K. A. & Trus, B. L. (1981) *Biosci. Rep.* 1:801
37. Almela, M., Maldonado, F., Otero, A. & Costa-Lopez, J. (1993) *J. Soc. Leather Technol. Chem.* 76:157
38. Gotsis, T., Spiccia, L. & Montgomery, K. C. (1993) *J. Soc. Leather Technol. Chem.* 76:195
39. Beek, J. & Sookne, A. M. (1939) *J. Am. Leather Chem. Assoc.* 34:641
40. Rose, G., Geselowitz, A., Lesser, G., Lee, R. & Zehfus, M. (1985) *Science* 229:834
41. Scholnick, F., Laio, L. L., Brown, E. M. & Fairheller, S. H. (1992) *J. Am. Leather Chem. Assoc.* 87:333
42. Tembe, B. L. & McCammon, J. A. (1984) *Comput. Chem.* 8:281
43. Lybrand, T. P., McCammon, J. A. & Wipff, G. (1986) *Proc. Natl. Acad. Sci. USA* 83:833
44. McCammon, J. A. (1987) *Science* 238:486

TABLE I

Portions of the amino acid sequence of Type I Collagen used for the microfibril fragment*

C1				C2				C3				C4				C5			
T	1	2	1'	T	1	2	1'	T	1	2	1'	T	1	2	1'	T	1	2	1'
17	GKN	GKA	GKN	329	GPp	GPA	GPp	251	GPA	GFV	GPA	173	GPR	GSR	GPR	95	GAA	GPS	GAA
18	GDD	GED	GDD	330	GPR	GIR	GPR	252	GEK	GEp	GEK	174	GAN	GPS	GAN	96	GEE	GEE	GEE
19	GEA	GHp	GEA	331	GRT	GSQ	GRT	253	GAp	GPS	GAp	175	GAp	GPp	GAp	97	GKR	GKR	GKR
20	GKP	GKP	GKP	332	GDA	GSQ	GDA	254	GAD	GEp	GAD	176	GND	GPD	GND	98	GAR	GST	GAR
21	GRp	GRp	GRp	333	GPA	GPA	GPA	255	GPA	GTA	GPA	177	GAK	GNK	GAK	99	GEp	GEI	GEp
22	GER	GER	GER	334	GPp	GPp	GPp	256	GAp	GPp	GAp	178	GDA	GEp	GDA	100	GPS	GPA	GPS
23	GPp	GVP	GPp	335	GPp	GPp	GPp	257	GTP	GTT	GTP	179	GAp	GVV	GAp	101	GLp	GPp	GLp
24	GPQ	GPQ	GPQ	336	GPp	GPp	GPp	258	GPQ	GPQ	GPQ	180	GAp	GAp	GAp	102	GPp	GPp	GPp
25	GAR	GAR	GAR	337	GPp	GPp	GPp	259	GIA	GLL	GIA	181	GSQ	GTA	GSQ	103	GER	GLR	GER
26	GLp	GFP	GLp	338	GPP	GPP	GPP	260	GQR	GAp	GQR	182	GAp	GPA	GAp	104	GGp	GNp	GGp
27	GTA	GTP	GTA	0				261	GVV	GFL	GVV	183	GLQ	GPS	GLQ	105	GSR	GSR	GSR
28	GLp	GLp	GLp	0				262	GLp	GLp	GLp	184	GMP	GIp	GMP	106	GFP	GLp	GFP

* C1-C5 refer to the individual triple helical segments used in the microfibril model.

The numbers 1 and 1' are used to distinguish the two alpha-1 chains in the triple helical portion of Type I Collagen.

Numbered tripeptides (T) are taken from the triple helical portion of Type I Collagen, which is composed of tripeptides 1-338.

The amino acids are represented by their one letter codes; P and p are proline and hydroxyproline, respectively.